An Enhanced Mining of Negative Association using Extended Patterns

SYNOPSIS

Submitted to

SWAMI RAMANAND TEERTH MARTHWADA UNIVERSITY, NANDED

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Subject of

Computer Science

Under the Faculty of Science and Technology

Submitted by

Mrs. Drishti Jaikumar Gangaramani

Under the Supervision of

Dr Renuka R. Londhe

Assistant Professor and HoD,

Rajarshi Shahu Mahavidyalaya (Autonomous), Latur

Research Centre

College of Computer Science and IT, Latur 413512

April 2022

Table of Contents

Sr. No.	Contents	Page No.
1	Introduction	1
2	Review of Literature	2
3	Objectives of Research	6
4	Methodology to be used	6
5	Importance of the study i. International status/National status ii. Significance of the study	7
6	Proposed work plan	8
7	References	9

An Enhanced Mining of Negative Association using Extended Patterns

Abstract of proposed work plan/ problem

Pattern mining concentrates on identifying rules that describe specific patterns within the data. Discovering patterns that appear many times in large input datasets is a well-known problem in data mining . Many algorithms, such as frequent item set mining, sequential pattern mining, and graph pattern mining, aim to capture frequent patterns. Extracting frequent patterns from databases has always been an imperative task for the data mining community. Literature has endowed plentiful endeavours to this research area with significant breakthroughs. Mining of extended patterns although being subsided has proved to be of vital importance in many domains. The application of extended pattern mining is inevitable and thus has become an emerging field of research. However, discovering extended patterns from databases comes up with numerous challenges. After mining the extended patterns the next issue is to extract the associations which exist among those patterns .There is lots of importance attached to the associations as most of the business decisions are dependent on the kinds of associations existing among the patterns. The associations could be either positive or negative. Most of the Research to date focussed on finding the positive associations among frequent and regular patterns and ignored the issue of negative associations, Also Businesses maintain their transactional data in several forms such as fixed database, Incremental database, Spatial database, Temporal database and also imbibe their decision by making use of data streams of data that flow across it .So Finding the negative association among those extended patterns considering different data sources is the main focus of this research.

1. Introduction

Data mining, sometimes used synonymously with "knowledge discovery," is the process of selecting large volumes of data for correlations, patterns, and trends. Various researches has been carried out to find out frequent patterns- a pattern that occurs frequently in a data set. Data Mining is a process of discovering various models, summaries, and derived values from a given collection of data. Frequent pattern mining in a large database is one of the interesting research areas in data mining. Frequent pattern generation basically depends on the associability of the items (attributes) in the database [1].It is being observed hardly any study has been conducted on Extended patterns.

The different patterns which comes under this category are rare patterns, uncertain patterns, compressed or approximate negative patterns, colossal patterns, high dimensional etc.

Association rule mining is one of the most popular data mining techniques to find associations among items in a set by mining necessary patterns in a large database. Typical association rules consider only items enumerated in transactions. Such rules are referred to as positive association rules. Negative association rules also consider the same items, but in addition consider negated items (i.e. absent from transactions). In recent years, there has been a significant research focused on finding interesting infrequent item sets leading to the discovery of negative association rules (NARs).The discovery of infrequent item sets is far more difficult than their counterparts, that is, frequent item sets. These problems include infrequent item sets discovery and generation of accurate NARs, and their huge number as compared with positive association rules. In medical science, for example, one is interested in factors which can either adjudicate the presence of a disease or write-off of its possibility. The vivid positive symptoms are often obvious; however, negative symptoms are subtler and more difficult to recognize and diagnose [2]. Negative association rules are useful in market-basket analysis to identify products that conflict with each other or products that complement each other [3].

2. Review of Literature and development in the subject (previous work done in the relevant area)

Various implementations have been introduced in the literature using various techniques.

Mr. Mukesh Poundekar and his co-workers [4] proposed a novel method for optimization of association rule mining. The algorithm which is used is combination of MLMS and min-max algorithm conducted the real world dataset such as heart disease data and some standard data used from UCI machine learning repository.

Ruchi Bhargava and Prof. Shrikant Lade [5] presented a new to generate all positive and negative class association rules using improved FP-tree and to build an accurate classification. In future this algorithm is performed for various large dataset and measurements will be taken.

Idheba Mohamad Ali O. Swesi and his colleagues [6] proposed model which shows an integration between two algorithms, the Positive Negative Association Rule (PNAR) algorithm and the Interesting Multiple Level Minimum Supports (IMLMS) algorithm, to propose a new approach (PNAR_IMLMS) for mining both negative and positive association rules from the interesting frequent and infrequent item sets mined by the IMLMS model. For future VARCC is applied that avoids generating uninteresting rules that may be discovered when mining positive and negative association rules.

Weimin Ouyang [7] proposed an efficient one-pass algorithm, called MPNAR-SW, for mining positive and negative Association rules over online data streams with a sliding window.

E. Bala Krishna, B. Rama, A. Nagaraju [8] proposed a new Improved FP tree (IFP– Tree) and a related algorithm (FISM) to mine negative association rules. The proposed novel approach produced useful and valid negative association

rules. Our approach is definitely capable than the existing FP–Growth based algorithms. In the future, we will expand the negative association rule mining tools.

Lieliang Zhang and his colleagues [9] presented an effective and efficient clustering algorithm FDLM for mining arbitrary shapes in large-scale and multi-dimensional database. Experiments on various databases are conducted, demonstrating the effectiveness and efficiency of our FDLM algorithm. Future work tells to extend FDLM algorithm to high-dimensional database to make a further performance analysis on it.

NVS Pavan Kumar and his co-workers [10] presented various algorithms presented using which mining of negative association that exist among regular and frequent patterns mined presented considering both local and Global Locations. Further, this aspect extended with hashing and indexing techniques on the database and additional parameters to strengthen the framework.

Thanh-Long Nguyen and his colleagues [11] proposed the CP-Miner algorithm for mining colossal patterns. CP-Miner mines colossal patterns using the CP-tree and pre-processing techniques to reduce the search space. VBUC uses bit vectors to represent item sets, when the number of items in the database is large, future work indicates how to compress this type of databases in the tree.

T Ramakrishnudu and R B V Sbramanyam [12] designed a new tree structure to store both the frequent and infrequent item sets for mining both positive and negative association rules. In the proposed method the database is scanned only once for mining positive and negative association rules, so it reduces the number of I/O operations.

Sikha Bagui and Probal Chandra Dhar [13] presented a Hadoop implementation of the A-priori algorithm to mine positive as well as negative association rules. So, for large datasets, a larger amount of parallelization that

Is more number of slave nodes with a higher block size is used which is more efficient.

Wannasiri Thurachon and Worapoj Kreesuradej [14] improved the efficiency of incremental frequent pattern mining with our fast incremental updating frequent pattern growth algorithm (FIUFP) and designed a new incremental conditional pattern tree (ICP-tree) suitable for incremental mining of frequent item sets.

D.S. Rajput, R.S. Thakur, G.S. Thakur [15] have used concept of fuzzy approach, in uncertain dataset and used multiple minimum supports to find fuzzy association rules in a given uncertain data set. Several investigating ways are derived to further reduce the data structure problems, and generated frequent patterns which will be useful for clustering.

Nikky Suryawanshi, Susheel Jain, Anurag Jain [16] presented a review of association rule mining with different approach and discuss advantage and demerits of rule mining. Also the generation process of negative and positive association rule mining was discussed. Future work indicates the optimization of negative rule generation technique using genetic algorithm.

E. Balakrishna, B. Rama and A. Nagaraju [17] presented the performance of A-priori, FP-Growth, and FISM and is compared by considering sample transactional database. Later it was found that FISM approach is efficient than A-priori and FP-Growth algorithm. Future directions indicates the development of the NAR mining tool.

E. Balakrishna, B. Rama, N. Satyavathi [18] developed FP-Tree with the linkage table to analyse the data with hash value. The proposed CFPT mine the positive and negative associative rule from the database. The CCM measure is

applied in the item sets to find the correlation. The result shows that the execution time of CFPT technique is low compared to the infrequent pattern mining. The future work of the method involves applying the proposed CFPT method for the uncertain data.

Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya [19] proposed a novel method for optimization of association rule mining. In this the proposed algorithm is combination of Interesting Multiple Level Minimum Support (IMLMS) and Genetic Algorithm (GA).

Sadeq Darrab, David Broneske, and Gunter Saake[20] present variations of the task of rare item set mining to discuss limitations of traditional rare item set mining algorithms. In the last, research opportunities and challenges for rare item set mining for future research are pointed out.

G.Sophana and her colleagues [21] showed study of various existing infrequent pattern mining techniques like Positive and Negative Association Rule (PNAR), Minimal Infrequent Item set Mining (MIIM), Rare Association Rule (RAR), Pattern-Growth Paradigm and Residual Trees, Optimization Rule Based Mining.

3. Objectives of Research

The main objective of this research is to find out infrequent or rare patterns (patterns that occur rarely) or negative patterns (items which are negatively correlated with each other)

1. To mine different extended patterns from various databases.
2. To extract the association rules by applying some algorithm.
3. To find out negative associations from the different extended patterns.

4. Methodology to be adopted:

Different data sources to be used for this research work

Many different types of data sources exist in nature that includes transactional, Spatial, temporal, Spatial-Temporal text, Multi-media, distributed, fixed database, Incremental database ,high dimensional database etc.Different mining techniques can be applied to these different kinds of database.

Applying the training dataset on these extended patterns, finding out frequent and infrequent item sets and extracting various association rules from them by applying some techniques.

The proposed research work will be carried out through following steps for mining negative association from extended patterns

a) Set of task relevant data to be mined. (database)
b) Kind of knowledge to be mined. (Association)
c) Certain knowledge to be used in discovery process.
d) Interestingness measures and thresholds for pattern evaluation.
e) Representation for visualizing the discovered patterns. (Rules)

The efforts will be made to increase the performance of proposed method.

The above mining process may need to use more terminology for getting accurate results.

5. Importance of the study

i. International status /National status:

The leading training and research institutes including "Harvard University, Stanford University, University of California, University of Cambridge are known for their excellence in conducting research in data mining domain.

Charu C. Aggarwal, Yorktown Heights New York -It states that negative pattern mining requires the use of bit-symmetric measures that treat the presence or absence of an item evenly. The traditional support-confidence measure is not designed for finding such patterns. Measures such as the statistical coefficient of correlation, $\chi 2$ measure, and collective strength are better suited for finding such positive or negative correlations between items.

Dr. Liu Bing ,National University of Singapore-With electronic commerce, a huge amount of information can be generated and collected, e.g., information about how customers search for products, what they buy, how they make purchase decisions, etc. Such information can be analysed and mined to discover interesting patterns

Azmi Ahmad, Faizah Shaari amd Awang L. Zalizah (2015), New York, United States -This paper discusses on finding meaningful outliers by hybridizing Rough Outlier Algorithm with Negative Association Rules (NAR).

ii. Significance of the study:

The discovery of interesting patterns can lead to important marketing and managerial decisions. All traditional association rule mining were developed to find only positive associations what about negative associations for example if someone wants to find out "Customers who want to buy coke will also buy Pepsi". In addition to positive associations, negative associations also provide valuable information in devising various marketing strategies.

There are also other areas where negative association mining can be used:

Medical diagnosis: In medical diagnosis, mammogram images are often used for cancer detection. In the entire image, however, only a small fragment indicates the cancerous pixels.

Hardware fault detection: Faults in hardware equipment's occur very rarely but need to be considered for the detection of equipment failure.

Credit card fraud detection: The credit card transactions stored in databases are millions in number. However, fraudulent activities in such transactions are rare and hence are very few in number.

6. Proposed work plan

Plan of work and targets to be achieved in first year:

During the first year an effort will be made to mine the data with the database and find out the extended patterns with the help of datamining approaches. Also, efforts will be made to study the existing algorithms relevant to our data. Two progress reports will be submitted at the end of this year.

Plan of work and targets to be achieved in second year:

During second year by making use of certain algorithms along with database positive and negative association rules will be found out and much focus will be given to mine the negative associations with our extended patterns .The research findings will be submitted in the form of two or more research papers to the internal research journals. Two progress reports will be submitted at the end of this year.

Plan of work and targets to be achieved in third year:

During this year, thesis writing will be started and publication of two research papers in reputed journals will be undertaken. Also, the research work will be presented in National/International seminars and conferences.

References:

1. Subrata Bose, Subrata Datta, “Frequent Pattern Generation in Association Rule Mining using Weighted Support” Department of Computer Science & Engineering NITMAS Kolkata, West Bengal, India, IEEE 2015.

2. Sajid Mahmood, Muhammad Shahbaz and Aziz Guergachi “Negative and Positive Association Rules Mining from Text Using Frequent and Infrequent Item sets” The Scientific World Journal, Volume 2014.

3. R.SUMALATHA,"MINING POSITIVE AND NEGATIVE ASSOCIATION RULES”, International Journal on Computer Science and Engineering, Vol.02, No.09, 2010.

4. Mr. Mukesh Poundekar, Mr.Amitkumar S Manekar,Prof.Mukesh Baghel,Prof. Hitesh Gupta, "Mining strong valid Association Rule form Frequent Pattern and Infrequent Pattern Based on Min-Max Sinc Constraints",IEEE 2014.

5. Ruchi Bhargava, Prof. Shrikant Lade "Effective Positive Negative Association Rule Mining Using Improved Frequent Pattern”, International Journal of Modern Engineering Research (IJMER), Vol.3, Issue.2, March-April 2013 pp-1256-1262

6. Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir,"Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets", IEEE 2012.

7. Weimin Ouyang,"Mining Positive and Negative Association Rules in Data Streams with a Sliding Window", IEEE 2013.

8. E. Bala Krishna,B.Rama,A.Nagaraju, "Mining of Negative Association Rules using Improved Frequent Pattern Tree",Vaagdevi College of Engineering, India.

9. Lieliang Zhang,Zhiyang Li, Weijiang Liu, Wenyu Quy,Yinan Wuz,"An efficient density-based clustering for multi-dimensional database",IEEE 2017.

10. NVS Pavan Kumar, Dr.JKR Sastry, Dr.K Raja Sekhara Rao, “Mining Distributed Databases for Negative Associations from Regular and Frequent Patterns “International Journal of Advanced Trends in Computer Science and Engineering, July-August 2019.

11. Thanh-Long Nguyen, Bay VO, Vaclav Snasel,"Efficient Algorithms for Mining Colossal Patterns in High Dimensional Databases", Knowledge-Based Systems (2017).
12. T Ramakrishnudu, R B V Sbramanyam, "Mining Positive and Negative Association Rules Using FII-Tree", International Journal of Advanced Computer Science and Applications, Vol.4, No.9, 2013.
13. Sikha Bagui and Probal Chandra Dhar, "Positive and negative association rule mining in Hadoop's MapReduce environment", Springer, Journal of Big Data, 2019.
14. WANNASIRI THURACHON AND WORAPOJ KREESURADEJ, "Incremental Association Rule Mining with a Fast Incremental Updating Frequent Pattern Growth Algorithm, IEEE 2021.
15. D.S.Rajput, R.S.Thakur, G.S.Thakur, "Fuzzy Association Rule Mining based Frequent Pattern Extraction from Uncertain Data", IEEE 2012.
16. Nikky Suryawanshi, Susheel Jain, Anurag Jain "A Review of Negative and Positive Association rule mining with multiple constraint and correlation factor", International Journal of Emerging Technology and Advanced Engineering, 2012.
17. E. Balakrishna, B.Rama and A. Nagaraju "Performance of Negative Association Rule Mining Using Improved Frequent Pattern Tree", Springer Nature Singapore Pte Ltd. 2018.
18. E. Balakrishna, B.Rama, N.Satyavathi "Mining of negative association rules using correlate frequent pattern tree", Emerging Trends in Materials Science, Technology and Engineering, 2021.
19. Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya "Mining Positive and Negative Association Rules from Frequent and Infrequent Pattern Using Improved Genetic Algorithm", IEEE 2013.
20. Sadeq Darrab, David Broneske and Gunter Saake, "Modern Applications and Challenges for Rare Itemset Mining "International Journal of Machine Learning and Computing, Vol. 11, No. 3, May 2021.

21. G.Sophana, Dr.V.Joseph Peter "Infrequent Pattern Mining Techniques: A Review" International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 4, April 2018.

Date: 18/04/2022

Place: Latur

(Drishti Gangaramani)	**(Dr Renuka R. Londhe)**
Research Scholar	Research Supervisor / Guide